AD-779 291

THE ENTROPY OF A MARKOV INFORMATION SOURCE

S. Kullback

George Washington University

Prepared for:

Office of Naval Research

19 April 1974

UNCLASSIFIED
Security Classification

AD-779291

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) THE GEORGE WASHINGTON UNIVERSITY DEPARTMENT OF STATISTICS WASHINGTON, D.C. 20006	2a. REPORT SECURITY CLASSIFICATION 2b. GROUP

3. REPORT TITLE
THE ENTROPY OF A MARKOV INFORMATION SOURCE

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
TECHNICAL REPORT

5. AUTHOR(S) (First name, middle initial, last name)
S. KULLBACK

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 19, 1974	12	0

8a. CONTRACT OR GRANT NO. N00014-67-A-0214-0015 b. PROJECT NO. NR-042-267 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 18 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
Unilmited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	OFFICE OF NAVAL RESEARCH STATISTICS & PROBABILITY PROGRAM ARLINGTON, VIRGINIA 22217

13. ABSTRACT

For a first-order and m-th order Markov source with a finite alphabet there are derived various relations and limiting behavior of expressions for the entropy.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

UNCLASSIFIED
Security Classification

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
MARKOV SOURCE						
ENTROPY STATIONARY DISTRIBUTION						

DD FORM 1 NOV 65 1473 (BACK)
(PAGE 2)

i-a

UNCLASSIFIED
Security Classification

THE ENTROPY OF A MARKOV INFORMATION SOURCE

by

S. KULLBACK

TECHNICAL REPORT NO. 18

APRIL 19, 1974

PREPARED UNDER CONTRACT N00014-67-A-0214-0015

(NR-042-267)

OFFICE OF NAVAL RESEARCH

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, D.C. 20006

S. Kullback

THE ENTROPY OF A MARKOV INFORMATION SOURCE

Let S be a first-order Markov source with alphabet $\{s_1, s_2, \ldots, s_q\}$, time-homogeneous transition probabilities $P(s_i \mid s_j)$ and stationary distribution $p_i = \mathrm{Prob}(s = s_i)$, $i = 1, 2, \ldots, q$. These define a simple stationary Markov chain with the <u>matrix of transition probabilities</u>

$$(1) \quad \underline{P} = \begin{pmatrix} P(s_1 \mid s_1) & P(s_2 \mid s_1) & \cdots & P(s_q \mid s_1) \\ P(s_1 \mid s_2) & P(s_2 \mid s_2) & \cdots & P(s_q \mid s_2) \\ \cdot & \cdot & \cdot & \cdot \\ P(s_1 \mid s_q) & P(s_2 \mid s_q) & \cdots & P(s_q \mid s_q) \end{pmatrix}$$

where

$$(2) \quad P(s_1 \mid s_j) + P(s_2 \mid s_j) + \ldots + P(s_q \mid s_j) = 1, \quad j = 1, 2, \ldots, q$$

and

$$(3) \quad p_j = p_1 P(s_j \mid s_1) + p_2 P(s_j \mid s_2) + \ldots + p_q P(s_j \mid s_q), \quad j = 1, 2, \ldots, q.$$

If we set $\underline{p}' = (p_1, p_2, \ldots, p_q)$ then in matrix notation the relations are expressed by

$$(4) \quad \underline{p}'\underline{P} = \underline{p}'.$$

If the source is in state s_i then its transitions to the different states s_j, $j = 1, 2, \ldots, q$ form a finite scheme

$$(5)\qquad \begin{matrix} s_1 & s_2 & \cdots & s_q \\ P(s_1|s_i) & P(s_2|s_i) & \cdots & P(s_q|s_i) \end{matrix}, \quad P(s_1|s_i) + \ldots + P(s_q|s_i) = 1 .$$

The entropy of the finite scheme in (5) we write as

$$(6)\qquad H(S|s_i) = -\sum_{j=1}^{q} P(s_j|s_i) \log P(s_j|s_i)$$

and may be regarded as a <u>measure of the amount of information obtained when the source (Markov process) advances one step forward, starting from the state s_i</u>. The mean value of the quantity in (6) over all states s_i, that is,

$$(7)\qquad H(S|S) = \sum_{i=1}^{q} p_i \, H(S|s_i) = -\sum_{i=1}^{q} \sum_{j=1}^{q} p_i P(s_j|s_i) \log P(s_j|s_i)$$

$$= -\sum_{i=1}^{q} \sum_{j=1}^{q} P(s_i, s_j) \log \frac{P(s_i, s_j)}{p_i}$$

$$= -\sum_{i=1}^{q} \sum_{j=1}^{q} P(s_i, s_j) \log P(s_i, s_j) + \sum_{i=1}^{q} \sum_{j=1}^{q} P(s_i, s_j) \log p_i$$

$$= -\sum_{i=1}^{q} \sum_{j=1}^{q} P(s_i, s_j) \log P(s_i, s_j) + \sum_{i=1}^{q} p_i \log p_i$$

$$= H(S \times S) - H(S) = H(S^2) - H(S)$$

may be regarded as <u>a measure of the mean amount of information obtained when the source (Markov process) moves one step ahead</u>. The quantity $H(S|S)$ which we shall call the <u>entropy</u> of the source obviously characterizes the source,

and is uniquely determined by the absolute probabilities p_i and the conditional probabilities $P(s_j|s_i)$, $1 \leq i \leq q$, $1 \leq j \leq q$.

Note that (7) may be written as

$$(8) \quad H(S^2) = H(S) + H(S|S),$$

which is essentially a special case of the general result

$$(9) \quad H(\chi,\psi) = H(\chi) + H(\psi|\chi).$$

It had been shown that generally

$$(10) \quad H(\psi) \geq H(\psi|\chi)$$

which in this case of a Markov source becomes

$$(11) \quad \log q \geq H(S) \geq H(S|S).$$

If we now consider sequences of $(n + 1)$ successive signals which we may also consider as constituted of the pair consisting of a sequence of n signals and a single signal then we have

$$(12) \quad H(S^{n+1}) = H(S^n,S) = H(S^n) + H(S|S^n).$$

But the Markov chain property that the conditional probability of a state depends only on the immediately preceding state

implies that

(13) $H(S|S^n) = H(S|S)$

or

(14) $H(S^{n+1}) = H(S^n) + H(S|S)$

and by successive application of (14) we get

(15) $H(s^{n+1}) = H(S) + nH(S|S)$, that is

(16) $H\,[(n+1)\text{-}\,\text{tuple}] = H(\text{single}) + n\,H(\text{transition})$.

We may also write

(17) $H(S^{n+1}) = H(S,S^n) = H(S) + H(S^n|S)$

so that comparing (15) and (17), we see that

(18) $H(S^n|S) = n\,H(S|S)$

and

(19) $H(S^{m+n}|S) = (m+n)H(S|S) = m\,H(S|S) + nH(S|S) =$

$= H(S^m|S) + H(S^n|S)$

that is, the entropy in (m + n) transitions is the sum of the

entropy in m and n transitions.

From (15) we see that

$$(20) \quad \frac{1}{n+1} H(S^{n+1}) = \frac{1}{n+1} H(S) + \frac{n}{n+1} H(S|S)$$

so that

$$(21) \quad \lim_{n \to \infty} \frac{1}{n+1} H(S^{n+1}) = H(S|S),$$

that is, the mean entropy per signal in a long sequence of signals is simply the entropy of the Markov source.

For a source without memory (independence) we have that

$$(22) \quad H(S^2) = 2H(S)$$

and for a Markov (order 1) source we have that

$$(23) \quad H(S^2) = H(S) + H(S|S).$$

Since we had in (11) that $H(S) \geq H(S|S)$, a measure of the correlation between successive signals may be taken as

$$(24) \quad 2H(S) - H(S) - H(S|S) = H(S) - H(S|S) = 2H(S) - H(S^2).$$

Note that

$$(25) \quad I = \sum_i \sum_j P(s_i, s_j) \log \frac{P(s_i, s_j)}{p_i\, p_j} \geq \left(\sum_i \sum_j P(s_i, s_j)\right) \log \frac{\sum_i \sum_j P(s_i, s_j)}{\sum_i \sum_j p_i\, p_j} = 0$$

and

$$(26) \quad I = \sum_i \sum_j P(s_i, s_j) \log P(s_i, s_j) - \sum_i \sum_j P(s_i, s_j) \log p_i - \sum_i \sum_j P(s_i, s_j) \log p_j$$

$$= \sum_i \sum_j P(s_i, s_j) \log P(s_i, s_j) - \sum_i p_i \log p_i - \sum_j p_j \log p_j$$

$$= - H(S^2) + 2H(S) = H(S) - H(S|S),$$

where $I = 0 \Leftrightarrow P(s_i, s_j) = p_i p_j$, that is, no memory.

The reader is reminded that all the preceding was relative to a first-order Markov source.

Let us now turn our attention to the case of an m-th order Markov source, that is, the conditional probability of a signal value or state depends only on the preceding m signal values or states. As before we have the relation in (12) (which we repeat here)

$$(12) \quad H(S^{n+1}) = H(S^n, S) = H(S^n) + H(S|S^n)$$

but now the specification of an m-th order source implies that

$$(27) \quad H(S|S^n) = H(S|S^m) \qquad n \geq m$$

or

(28) $H(S^{n+1}) = H(S^{n}) + H(S|S^{m})$

and by successive application of (28) we get

(29) $H(S^{n+1}) = H(S^{m}) + (n - m + 1)\, H(S|S^{m}), \quad n \geq m.$

As in (11) we may also write

(30) $H(S^{n+1}) = H(S^{m}, S^{n-m+1}) = H(S^{m}) + H(S^{n-m+1}|S^{m})$

so that comparing (29) and (30) we see that

(31) $H(S^{n-m+1}|S^{m}) = (n - m + 1)\, H(S|S^{m}), \qquad n \geq m,$

and

(32) $H(S^{n_1+n_2}|S^{m}) = (n_1 + n_2) H(S|S^{m}) = n_1 H(S|S^{m}) + n_2 H(S|S^{m})$

$$= H(S^{n_1}|S^{m}) + H(S^{n_2}|S^{m}), \quad n_1 \geq m, \; n_2 \geq m.$$

From (29) we see that

(33) $\dfrac{1}{n+1} H(S^{n+1}) = \dfrac{1}{n+1} H(S^{m}) + \dfrac{n-m+1}{n+1} H(S|S^{m}), \qquad n \geq m$

so that

(34) $\lim\limits_{n \to \infty} \dfrac{1}{n+1} H(S^{n+1}) = H(S|S^{m}),$

that is, the mean entropy per signal in a long sequence of signals is simply the entropy of the m-th order Markov source.

If we use the notation $P(S^n)$ to represent the probability of a sequence of n successive signals, then by the convexity property

(35) $$I = \sum_{S^m}\sum_{S} P(S^m,S) \log \frac{P(S^m,S)}{P(S^m)P(S)} \geq \sum_{S^m}\sum_{S} P(S^m,S) \log \frac{\sum_{S^m}\sum_{S} P(S^m,S)}{\sum_{S^m} P(S^m)\sum_{S} P(S)}$$

$$= 1 \log \frac{1}{1} = 0, \text{ or}$$

(36) $$\sum_{S^m}\sum_{S} P(S^m,S) \log P(S^m,S) - \sum_{S^m}\sum_{S} P(S^m,S)\log P(S^m) - \sum_{S^m}\sum_{S} P(S^m,S)\log P(S)$$

$$= \sum_{S^{m+1}} P(S^{m+1}) \log P(S^{m+1}) - \sum_{S^m} P(S^m) \log P(S^m) - \sum_{S} P(S) \log P(S)$$

$$= - H(S^{m+1}) + H(S^m) + H(S) = - H(S^m) - H(S|S^m) + H(S^m) + H(S)$$

so that

(37) $$I = H(S) - H(S|S^m) \geq 0$$

with $I = 0 \Leftrightarrow$ a signal is independent of the preceding m signals, is a measure of the relation between a signal and the preceding m signals in an m-th order Markov source.

To assist the reader to relate the exposition and notation in these notes with that in the text by Abramson we indicate equivalent values and results in the following table

Abramson	Kullback	
$H(S)$	$H(S \mid S)$	order 1 source
$H(S)$	$H(S \mid S^m)$	order m source
$H(\overline{S})$	$H(S)$	
$H(\overline{S^n})$	$H(S^n)$	
$H(S^n)$	$H(S^n \mid S)$	order 1 source
$H(S^n)$	$H(S^n \mid S^m)$	order m source
(2-29) p. 28	(11)	
(2-37) p. 31	(18)	
(2-41) p. 31, (2-40) p. 31	(15)	
(2-42) p. 31	(29)	
(2-45) p. 32	(34)	
(2-44) p. 32	(10) with $\mu=S^n$, $\chi=S^m$, order m source	

cf. On the entropy of Markov chains by G.A. Ambarcumjan
<u>Izv. Akad. Nauk. Armjan. SSR Ser. Fiz. Mat. Nauk.</u> 11(1958) no. [illegible],
31-40. <u>Selected Papers in Math. Statist. and Probability</u>,
Vol. 4 (1963), pp. 1-11.

EXAMPLE - Entropy Markov Chain

	s_1	s_2	s_3	s_4
s_1	.2	.8	0	0
s_2	0	0	.1	.9
s_3	0	0	.2	.8
s_4	.7	.3	0	0

Stationary Probabilities

$$.2p_1 + .7p_4 = p_1$$
$$.8p_1 + .3p_4 = p_2$$
$$.1p_2 + .2p_3 = p_3$$
$$.9p_2 + .8p_3 = p_4$$

$$-8p_1 + 7p_4 = 0$$
$$p_2 - 8p_3 = 0$$
$$72p_3 + 8p_3 = 10p_4$$
$$8p_3 = p_4 = p_2$$
$$7/8p_4 + p_4 + 1/8p_4 + p_4 = 1$$
$$p_4 = 1/3, p_1 = 7/24$$
$$p_2 = 1/3, p_3 = 1/24$$

Using (7)

$$H(S|S) = 7/24H(.2,.8)+1/3H(.9,.1)$$
$$+1/24H(.2,.8) + 1/3H(.7,3)$$
$$=7/24(.721928 + 1/3(.468996)$$
$$+1/24(.721928)+1/3(.881291) \text{ bits}$$
$$= .690738 \text{ bits}$$

$$H(S) = 7/24 \log 24/7 + 1/3 \log 3 + 1/24 \log 24 + 1/3 \log 3$$
$$= 1/3 \log 24 + 2/3 \log 3 - 7/24 \log 7$$
$$= 1/3(4.584962) + 2/3(1.584962) - 7/24(2.807355) \text{ bits}$$
$$= 1.76615 \text{ bits}$$

S^2	
$S_1 S_1$	$.2(7/24) = 14/240$
$S_1 S_2$	$.8(7/24) = 56/240$
$S_1 S_3$	0
$S_1 S_4$	0
$S_2 S_1$	0
$S_2 S_2$	0
$S_2 S_3$	$.1(1/3) = 8/240$
$S_2 S_4$	$.9(1/3) = 72/240$
$S_3 S_1$	0
$S_3 S_2$	0
$S_3 S_3$	$.2(1/24) = 2/240$
$S_3 S_4$	$.8(1/24) = 8/240$
$S_4 S_1$	$.7(1/3) = 56/240$
$S_4 S_2$	$.3(1/3) = 24/240$
$S_4 S_3$	0
$S_4 S_4$	0

m	$n \log_2 n$
14	53.30297
56	325.2119
8	24.0000
72	444.2346
2	2.0000
24	110.0391

$2H(S) - H(S^2) =$

$$\begin{array}{r} 3.53230 \\ -2.45689 \\ \hline 1.07541 \end{array}$$

$H(S) - H(S|S)$

$$= \begin{array}{r} 1.76615 \\ -\ .69074 \\ \hline 1.07541 \end{array}$$

$$H(S^2) = \frac{14}{240}\log\frac{240}{14} + \frac{56}{240}\log\frac{240}{56} + \frac{8}{240}\log\frac{240}{8} + \frac{72}{240}\log\frac{240}{72}$$

$$+ \frac{2}{240}\log\frac{240}{2} + \frac{8}{240}\log\frac{240}{8} + \frac{56}{240}\log\frac{240}{56} + \frac{24}{240}\log\frac{240}{24}$$

$$= \log 240 - \frac{1308.0005}{240}$$

$$= 5.321928 + 2.584962 - 5.45 \text{ bits} = 2.45689 \text{ bits}$$

$$H(S^2) - H(S) = 2.45689 - 1.76615 = .69074 \text{ bits}$$